

Early EEG correlates of word frequency and contextual predictability in reading

Sara C. Sereno, Christopher J. Hand, Aisha Shahid, Ian G. Mackenzie & Hartmut Leuthold

To cite this article: Sara C. Sereno, Christopher J. Hand, Aisha Shahid, Ian G. Mackenzie & Hartmut Leuthold (2020) Early EEG correlates of word frequency and contextual predictability in reading, *Language, Cognition and Neuroscience*, 35:5, 625-640, DOI: [10.1080/23273798.2019.1580753](https://doi.org/10.1080/23273798.2019.1580753)

To link to this article: <https://doi.org/10.1080/23273798.2019.1580753>



© 2019 The Author(s). Published by Informa UK Limited, trading as Taylor & Francis Group



[View supplementary material](#)



Published online: 03 Apr 2019.



[Submit your article to this journal](#)



Article views: 1062



[View related articles](#)



[View Crossmark data](#)



Citing articles: 1 [View citing articles](#)

REGULAR ARTICLE



Early EEG correlates of word frequency and contextual predictability in reading

Sara C. Sereno^{a,b}, Christopher J. Hand^c, Aisha Shahid^{b,d}, Ian G. Mackenzie^e and Hartmut Leuthold^e

^aInstitute of Neuroscience and Psychology, University of Glasgow, Glasgow, UK; ^bSchool of Psychology, University of Glasgow, Glasgow, UK;

^cDepartment of Psychology, Glasgow Caledonian University, Glasgow, UK; ^dSchool of Psychology, The Open University, Edinburgh, UK;

^eDepartment of Psychology, University of Tübingen, Tübingen, Germany

ABSTRACT

Previous research into written language comprehension has been equivocal as to whether word frequency and contextual predictability effects share an early time course of processing. Target word frequency (low, high) and its predictability from prior context (low, high) were manipulated across two-sentence passages. Context sentences were presented in full, followed by word-by-word presentation (300 ms SOA) of target sentences. ERPs were analysed across left-to-right and anterior-to-posterior regions of interest within intervals from 50 to 550 ms post-stimulus. The onset of significant predictability effects (50–80 ms) preceded that of frequency (P1, 80–120 ms), while both main effects were generally sustained through the N400 (350–550 ms). Critically, the frequency-predictability interaction became significant in the P1 and was sustained through the N400, although the specific configuration of effects differed across components. The pattern of findings supports an early, chronometric locus of contextual predictability in recognising words during reading.

ARTICLE HISTORY

Received 5 February 2018

Accepted 2 January 2019

KEYWORDS

Word frequency; contextual predictability; reading; ERPs

Introduction


Two fundamental variables affecting how fast a word is recognised during reading are its frequency of occurrence and its predictability from the prior context (Rayner & Sereno, 1994). The relative timing of perceptual and knowledge-based processing is significant in determining the underlying neurocircuitry of word recognition. While word frequency effects are thought to reflect early lexical processing (Sereno & Rayner, 2000, 2003), the temporal locus of contextual effects is less certain, both theoretically and empirically.

Modular models posit that context can only operate post-lexically (Fodor, 1983), while interactive models maintain that context can directly affect lexical processing (McClelland, 1987). The relative timing of bottom-up, perceptual and top-down, contextual processes has been used as support for these alternative models of word recognition (for a discussion, see Dambacher et al., 2012; Sternberg, 1969). Demonstrating an earlier (e.g. Dambacher, Rolfs, Göllner, Kliegl, & Jacobs, 2009) or later (e.g. Murray & Forster, 2004) time course of predictability effects would provide support for interactive or modular accounts, respectively. Specifically, an early interaction of frequency and predictability would provide strong evidence that lexical access

is guided by top-down processes (e.g. Dambacher et al., 2012).

Prior investigations of frequency-predictability effects have employed reaction time (RT), eye movement, and electrophysiological paradigms (see Hand, Miellet, O'Donnell, & Sereno, 2010, for a review). In general, a sentence context is provided and the response to a subsequent target word is measured. Target words are either commonly used, high frequency (HF) words or low frequency (LF) words which are used far less often. Sentence contexts are either biasing or not with respect to the target, making it more or less predictable. The findings are mixed. Behavioural studies using naming or lexical decision tasks generally find an interaction of frequency and predictability in RTs, with greater predictability effects in LF than HF words (Stanovich & West, 1983; West & Stanovich, 1982). In several eye movement reading studies, fixation time measures on targets have shown additive effects (Hand et al., 2010; Hand, O'Donnell, & Sereno, 2012; Kliegl, Grabner, Rolfs, & Engbert, 2004; Miellet, Sparrow, & Sereno, 2007; Rayner, Ashby, Pollatsek, & Reichle, 2004). However, some eye movement studies have reported interactive findings (launch site analysis in Hand et al., 2010; Inhoff, 1984; Kliegl, Nuthmann, & Engbert, 2006; Sereno, Hand, Shahid, Yao, & O'Donnell, 2018), with LF

CONTACT Sara C. Sereno  Sara.Sereno@glasgow.ac.uk

 Supplemental data for this article can be accessed <https://doi.org/10.1080/23273798.2019.1580753>.

© 2019 The Author(s). Published by Informa UK Limited, trading as Taylor & Francis Group

This is an Open Access article distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

words demonstrating greater predictability effects than HF words. Electrophysiological studies have found similar interactive effects of frequency and predictability expressed in voltage amplitudes of different components of event-related potentials (ERPs), including the N1 (132–192 ms post-stimulus; Sereno, Brewer, & O'Donnell, 2003) and N400 (300–500 ms post-stimulus; Dambacher, Kliegl, Hofmann, & Jacobs, 2006; van Petten & Kutas, 1990).

The present study utilised an electrophysiological approach to provide a temporally refined examination of the combined effects of frequency and predictability. It is therefore worthwhile to first describe in more detail previous relevant ERP findings. In studies presenting words in isolation, robust word frequency effects have been reported in the N1 time range (~130–200 ms post-stimulus), with LF words eliciting a greater amplitude than HF words over lateral posterior electrodes (e.g. Hauk & Pulvermüller, 2004; Scott, O'Donnell, Leuthold, & Sereno, 2009; Sereno et al., 2003; Sereno, Rayner, & Posner, 1998). However, there is also evidence for a frequency effect occurring even earlier, in the P1 (100–120 ms), with increasing ERP amplitudes over left-posterior electrodes associated with decreasing word frequency (Hauk, Davis, Ford, Pulvermüller, & Marslen-Wilson, 2006). In time intervals later than 300 ms, LF words elicited more negative-going amplitudes compared to HF words (Hauk & Pulvermüller, 2004; Rugg, 1990; but see Polich & Donchin, 1988, for a reverse effect in P300 amplitude).

Crucially, when studying word frequency effects in the ERP waveform at the sentence or discourse level, the predictability of the forthcoming critical word is an important factor that influences the processing of linguistic input. Predictability effects have been found in the ERP waveform as early as 50–90 ms after critical word onset during discourse comprehension (Dambacher et al., 2009). In the 200–300 ms time range, a larger ERP positivity (P200) over anterior midline sites has been found when incoming linguistic information was unexpected or affectively salient (Leuthold, Kunkel, Mackenzie, & Filik, 2015), in line with the view that the P200 component indicates the enhanced attentional processing of such visual input (e.g. Hillyard & Münte, 1984). After approximately 200 ms and over centroparietal midline sites, a negative-going ERP deflection typically starts to develop in the ERP waveform, peaking at about 400 ms after critical word onset. This N400 component is taken to reflect lexical-semantic processing – words that are unexpected or a poor fit with context at the sentence- or discourse-level elicit a larger N400 than those that are expected or a good fit (e.g. DeLong, Urbach, & Kutas, 2005; Kutas & Hillyard, 1984; van Berkum, Hagoort, & Brown, 1999; for a review, see Kutas & Federmeier, 2011).

Of key interest for present purposes are six ERP studies that have investigated the combined effects of word frequency and predictability. We present the studies in chronological order. In the first study, van Petten and Kutas (1990) examined word frequency effects dependent on the position of the critical word within the sentence. Sentences were presented word-by-word, with each word displayed centrally for 200 ms using a 900 ms stimulus onset asynchrony (SOA). They did not directly manipulate predictability, but instead assumed that words become more predictable the later they appear in sentences. They showed a larger N400 (350–500 ms) for LF than HF words occurring early in the sentence but not at intermediate or sentence-final positions. They suggested that this attenuation of frequency effects indicated that contextual constraint strongly influences lexical processing.

In the second study, Sereno et al. (2003) presented HF, LF, and ambiguous words as sentence-final targets within neutral or biasing contexts, using 32 items per condition. Their ambiguous words – having a strongly dominant sense and a much weaker subordinate sense – were functionally equivalent to HF words in neutral contexts and LF words in (subordinate-meaning) biasing contexts. Individual words from sentences were presented centrally for 225 ms with a 450 ms SOA (N.B. the sentence-final target was presented for 495 ms). In addition to a word frequency effect, they found an interaction of word type and context in the N1 (132–192 ms), with stronger frequency effects in neutral than biasing contexts. They suggested their findings supported an early, interactive account of lexical processing.

In the third study, Dambacher et al. (2006) examined word frequency and word position effects within sentences together with word predictability effects. Their targets comprised all open-class words from the 144-sentence Potsdam Sentence Corpus (PSC; Kliegl et al., 2004), with varying frequencies and predictabilities. Sentences were presented word-by-word, with each word displayed centrally for 250 ms with a 700 ms SOA. They found an ERP positivity (P2) between 140 and 200 ms over anterior electrodes that increased with decreasing word frequency and was independent from predictability. In the N400 (300–500 ms), amplitude increased with decreasing predictability. More importantly, in line with the findings of van Petten and Kutas (1990), the frequency effect in the N400 amplitude (i.e. a larger N400 for LF than HF words) decreased with increasing predictability. They suggested this result indicated that LF words can acquire greater benefit from contextual information.

In the fourth study, Penolazzi, Hauk, and Pulvermüller (2007) manipulated word frequency (LF, HF), predictability (low, high), as well as word length (four or six letters, on average) within single sentences, using 35 items per

condition. Sentences were presented word-by-word, with each word presented centrally for 300 ms, using a 700 ms SOA. Separate frequency and predictability effects occurred interactively with word length between 110 and 130 ms after critical word onset, presumably indicating early lexical access. In the subsequent 170–190 ms time range, there was a more negative-going ERP amplitude for low than high predictability conditions over midline-parietal electrodes. In the same time interval, LF words elicited more negative amplitudes than HF words over midline-central and -posterior electrodes, but only when word length was short. Lastly, N400 amplitude (250–450 ms), as anticipated, was larger for low than high predictability conditions over posterior sites. Although frequency and predictability each interacted with word length, in contrast to the prior studies, frequency and predictability did not interact with each other. Penolazzi et al. concluded that lexical access and semantic context integration are distinct systems but that both are influenced by lower-level orthographic and phonological processing.

In the fifth study, Dambacher et al. (2012) examined frequency (LF, HF) and predictability (low, high) effects on target words presented in the second sentence of two-sentence passages, with 36 items per condition. Passages were presented word-by-word, with each word displayed centrally for 250 ms. Word-to-word SOAs were manipulated across three separate experiments, using SOAs of 700, 490, and 280 ms. At the 700 and 490 ms SOAs, Dambacher et al. (2012) found significant main effects of both frequency and predictability in the P2 (240–300 ms) and significant effects of predictability in the N400 (300–500 ms), but no evidence of an interaction in any component (N.B. there was a marginal effect of frequency at the 490 ms SOA in the N1, from 140 to 210 ms). At the 280 ms SOA, using slightly different time windows for analysis, a different pattern emerged. Significant frequency effects were found in the N1 (190–260 ms) and P2 (300–360 ms) components, significant predictability effects were found in the P2 (300–360 ms) and N400 (300–500 ms) components, and a significant interaction was found in the early N1 (135–155 ms) component. Dambacher et al. (2012) suggested that presentation rate affects word recognition processes. Specifically, they proposed that their fastest SOA of 280 ms (i.e. the one most representative of a normal reading rate) allows for immediate top-down processing to constrain word identification. In a related study, Dambacher et al. (2009) presented the same materials using a 280 ms SOA and reported an earlier significant effect of predictability from 50 to 90 ms post-stimulus.

Finally, in the sixth study, Kretzschmar, Schlesewsky, and Staub (2015) investigated frequency (LF, HF) and

predictability (low, high) effects on target words presented within single sentences during normal reading. Although they used 40 items per condition, target words were repeated across levels of predictability. Eye movements and fixation-related potentials (FRPs) were simultaneously recorded. Earlier, Dimigen, Sommer, Hohlfeld, Jacobs, and Kliegl (2011) also implemented a co-registration paradigm. However, similar to Dambacher et al. (2006), Dimigen et al.'s materials comprised all open-class words from the PSC, with words categorised as having low, medium, or high predictability. In their FRP data, Dimigen et al. found that significant predictability effects (low vs. high) were limited to the N400 (300–500 ms) component. Kretzschmar et al. used consecutive analysis windows of 50 ms from 150 to 700 ms after fixation onset. They found significant sustained effects of predictability from 150 to 650 ms. However, the frequency effect only emerged in the 500–550 ms window and the frequency-predictability interaction was restricted to the 300–350 ms time window. Kretzschmar et al. suggested that the interaction between top-down and bottom-up processes is relatively delayed, with the N400 demonstrating particular sensitivity to such interplay.

Taken together, the pattern of frequency-predictability effects is quite mixed, not only across electrophysiological studies, but also in RT and eye movement reading studies (for a review, see Hand et al., 2010). However, there are paradigm-specific constraints as well as aspects of the stimulus materials, themselves, which may have influenced results and, hence, limit their generalisability. For example, measures having longer latencies (e.g. RTs occurring in the range of 500–800 ms; later-onset components of the ERP such as the N400) are less likely to solely reflect the immediate, automatic stages of word recognition and are more susceptible to influences from conscious, strategic processing (for a discussion, see Gold et al., 2006). Similarly, the use of word presentation rates in ERP studies (often 500–1000 ms/word) that are several times slower than the normal reading rate of ~200–300 ms/word also encourages non-automatic processing. In order to appropriately differentiate modular and interactive accounts, the use of measures sensitive to the early stages of lexical access is necessary.

In terms of materials, most studies to date have used fairly brief contexts, limited to a handful of words preceding the target within a single sentence. For example, the contexts of Penolazzi et al. (2007) comprised five words and often employed semantic associates or set phrases (e.g. *electrical power*, *shook with fear*; targets underlined). The use of condensed contexts increases the likelihood of employing semantic primes in order to make a

Table 1. Example materials from Dambacher et al. (2012) and Kretzschmar et al. (2015).

Condition		Dambacher et al. (2012) example experimental materials	
LF-LP	(a)	The man on the picture fiddled around with models of Columbus' fleet. In his right hand he held a <u>sceptre</u> of considerable length.	
	(b)	Johannes heard a huge plane approaching from some distance. He gazed intensely into the <u>tunnel</u> and listened carefully.	
LF-HP	(a)	The man on the picture wore a golden crown and sat stately on a throne. In his right hand he held a <u>sceptre</u> of considerable length.	
	(b)	Before walking through the mountain, Johannes wanted to make sure, that no train was approaching. He gazed intensely into the <u>tunnel</u> and listened carefully.	
HF-LP	(a)	The man on the picture wore a golden crown and sat stately on a throne. In his right hand he held a <u>ship</u> of considerable length.	
	(b)	Before walking through the mountain, Johannes wanted to make sure, that no train was approaching. He gazed intensely into the <u>sky</u> and listened carefully.	
HF-HP	(a)	The man on the picture fiddled around with models of Columbus' fleet. In his right hand he held a <u>ship</u> of considerable length.	
	(b)	Johannes heard a huge plane approaching from some distance. He gazed intensely into the <u>sky</u> and listened carefully.	
Condition		Kretzschmar et al. (2015) example experimental materials	
LF-LP		I want to go to graduate school so I can help people with <u>anorexia</u> recover from their illness.	
LF-HP		The extremely skinny model looked like she suffered from <u>anorexia</u> and a lack of sleep.	
HF-LP		Yesterday I noticed that we passed by a <u>church</u> on our way to the apartment.	
HF-HP		On Sunday morning, the nun went to pray at the <u>church</u> and then went for a walk.	

Note: Target words are underlined. The Dambacher et al. materials are translations of the original German and are reproduced with permission from their Figure 1 (p. 1856), copyright 2012, Elsevier. The Kretzschmar et al. materials are reproduced with permission from their Table 1 (p. 1651), copyright 2015, American Psychological Association. LF=low frequency; HF=high frequency; LP=low predictability; HP=high predictability.

target more predictable. While semantic associates facilitate target identification, when in close proximity to the target, such "context" is thought to originate at the lexical level and not from a higher discourse or message level (Forster, 1979). As such, intralexical priming effects cannot differentiate modular from interactive accounts of processing. Moreover, the use of longer texts preceding target words may allow for context effects to develop more fully (Hand et al., 2010). Another potential difficulty with materials in many ERP (as well as RT) studies is having the target word as the last word of the sentence (see Dimigen et al., 2011). For example, sentence-final words typically elicit more positive-going ERPs (Friedman, Simson, Ritter, & Rapin, 1975; Hagoort, 2003). Another aspect of ERP studies is that they require a relatively high number of items per condition. Most frequency-predictability studies (see above) have used 30–40 items per condition.

Finally, a variety of approaches have been adopted when creating materials in 2 (Frequency: LF, HF) × 2 (Predictability: low, high) designs. There are strengths and limitations of each approach utilised thus far. In Dambacher et al. (2012), materials comprised two-sentence passages, with an initial context sentence and a subsequent sentence containing an LF or HF target. For target sentences, an LF or HF word was identified in each of the 144 sentences of the PSC. An alternative target word of equal length but of opposing frequency was then selected for each sentence frame. A preceding, context sentence was then constructed – one for the LF and one for the HF target sentences – designed to increase the predictability of the upcoming target (i.e. the high predictability conditions). For low predictability

conditions, these initial context sentences were switched. Thus, for each of the 144 sentences of the PSC, four different passages were created – high predictability (HP) and low predictability (LP) versions containing either LF or HF targets. This method ensures that the local content of the target sentence is identical between LF and HF conditions. However, having the initial HP context sentence perform a dual role – that is, act as the LP context for the alternative target – sometimes introduced anomalies. In the examples presented in Table 1 (upper section), it would not be appropriate to characterise targets in LP versions as simply having a lower level of predictability. In these cases, the initial contexts make the targets semantically awkward if not anomalous. In Kretzschmar et al. (2015), LF and HF targets (40 of each) appeared in single sentences whose preceding contexts were constructed to be either LP or HP. For each target word, individualised contexts were created for each LF or HF word, as seen in the examples in Table 1 (lower section). The practice of using uni-functional contexts has the advantage that target words are genuinely of higher or lower predictability. In particular, in LP contexts, target words do not stand out as semantic misfits. The downside of using such contexts is that the local sentence content (i.e. the words immediately preceding and following the target) is different across frequency conditions.

The current experiment addressed these issues. LF and HF target words were presented in two-sentence passages, with the first sentence establishing a context and the target embedded within the second sentence. For each target, two different context sentences were constructed – one that was neutral, making the target

LP and one that was biasing toward the target, making the target HP. The second, target sentence was identical across context conditions. Care was taken to ensure that the pre-target region of the second sentence was relatively neutral and did not contain, for example, intralexical primes of the subsequent target. Two sets of materials were prepared. Each set was comprised of 62 items from each of the four possible conditions (LF-LP, LF-HP, HF-LP, and HF-HP). Each set of materials was presented to a distinct participant group, such that targets that were LP in one set were HP in the other and vice versa. In this way, all participants were presented with all targets, but in only one of the contexts to avoid repetition of target sentences. The electroencephalogram (EEG) was recorded while participants read the passages. The first, context sentence was presented in its entirety. This was followed by a word-by-word presentation of the second, target sentence. In order to more closely imitate the speed of normal reading, a presentation rate of 300 ms/word was used. With these methodological controls in place, the pattern of our results should elucidate the degree to which contextual factors can influence lexical processing.

While Dambacher et al. (2012) examined the frequency-predictability interaction using lengthy pre-target contexts and a presentation rate akin to normal reading, their practice of mixing and matching contexts and targets was potentially problematic (i.e. possible target anomalies). Kretzschmar et al. (2015) avoided this by using individualised contexts per target. However, their pre-target contexts were relatively short, the local sentence content across conditions differed, and target words were repeated across predictability conditions. We have built upon these investigations by employing the following methodology: (1) a lengthier, separate discourse context; (2) bespoke LP and HP context sentences that preceded an identical target sentence; (3) a target word embedded within the body of a sentence (not sentence-final); (4) a rapid, reading-like presentation rate; and (5) a high number of experimental items per condition. First, contextual information (LP or HP) was provided via an entire initial sentence, while the second target sentence remained relatively neutral in its pre-target region (see, e.g. Sereno et al., 2018). When materials comprise single-line sentences (sometimes using sentence-final targets), biasing contextual information tends to take the form of semantic associates appearing only a word or two away from the target, fostering intralexical priming and not higher-level discourse processes. Second, the LP context sentences were specifically created to be relatively neutral with respect to target words in target sentences. In other studies, HP contexts for LF and HF targets were switched in order to create corresponding LP contexts

(e.g. Dambacher et al., 2012; Rayner et al., 2004). This results in target words that are often semantically anomalous in context, not simply of lower predictability. As such, findings interpreted as (facilitatory) predictability effects may additionally include (inhibitory) anomaly effects. Third, target words were embedded within target sentences and were not sentence-final. Moreover, the same (second) target sentence was used in LP and HP conditions (only the first, context sentence differed). Thus, target position was not confounded with predictability, but was kept identical across context conditions. Finally, words in the target sentence were presented at a rapid, reading-like rate (300 ms SOA). Slower presentation rates (e.g. SOAs of 500–1000 ms) that have typically been used disrupt the normal flow of reading and encourage strategic processing.

Method

Participants

Thirty native English-speaking members of the University of Glasgow community (21 female; mean age 22) took part in the experiment. All had normal or corrected-to-normal vision, had not been diagnosed with any reading disorder, were right-handed, and were either paid £10 or given course credit for their participation. The study conformed to British Psychological Society ethical guidelines and protocols. Data from an additional 8 participants were collected but were excluded from the analyses due to excessive alpha activity ($N = 3$), substantial drifts in EEG activity at multiple electrodes ($N = 3$), or having less than 50% of trials per condition remaining after artifact correction and rejection ($N = 2$).

Design and materials

A 2 (Frequency: LF, HF) \times 2 (Predictability: LP, HP) design was used, with 62 target items in each of the four conditions. Target word length was limited to a range of 4–8 letters. Word frequencies were acquired from the British National Corpus (BNC), a corpus of 90 million written word tokens (Davies, 2004). Each LF or HF target appeared in the second (target) sentence of two-line passages. The average position of LF and HF targets was word 5.86 and 5.74 of the sentence, respectively. The first (context) sentence provided either a relatively neutral or semantically biasing context for each target, producing LP and HP conditions, respectively. The average lengths of first sentences were matched across conditions and were as follows: 61, 60, 61, and 59 characters for LF-LP, LF-HP, HF-LP, and HF-HP conditions, respectively. The level of target predictability was

Table 2. Specifications of target words.

	N	Length	Frequency	Cloze	
				LP	HP
LF	124	5 (.6)	8 (6)	.03 (.07)	.61 (.25)
HF	124	5 (.6)	99 (90)	.06 (.12)	.61 (.29)

Note: Condition means with standard deviations in parentheses. Units of measurement are as follows: Length in number of letters; Frequency in occurrences per million; Cloze probability on a scale of 0 (target word not guessed) to 1 (target word correctly guessed). LF = low frequency; HF = high frequency; LP = low predictability; HP = high predictability.

determined by a Cloze probability task administered to two groups of 20 participants (none of whom participated in the main experiment). These participants were given each item up to but not including the target word and were asked to generate the next word in the sentence. Items were scored as "1" for correct responses (i.e. the unseen target word) and "0" for all other guesses. Each participant group was only presented with one of the two context sentences for any target word; both groups were presented with equal numbers of neutral and biasing contexts. Target word specifications of length, frequency, and Cloze values¹ are presented in Table 2. All target words are listed in the Appendix (see Supplemental data). Example sets of materials for LF and HF targets in LP and HP conditions are presented in Table 3.

We conducted a small post-hoc rating study to examine the plausibility of targets within their contexts, for both the current set of materials and those of Dambacher et al. (2012). For both sets of materials, two lists were created, counterbalanced such that targets appeared only once per list, in either their LP or HP context.

Table 3. Example materials.

Condition	Passages comprised of context and target sentences
LF-LP	(a) Alison normally steamed her food but today she was in a hurry. She added the <u>onion</u> and peppers into the oil in the pan.
	(b) Jill shuddered as the rain battered against her doors and windows. In the morning, she noticed an enormous <u>stain</u> on the carpet.
LF-HP	(a) Alison's eyes were watering as she chopped the vegetables. She added the <u>onion</u> and peppers into the oil in the pan.
	(b) Jill's friends were drinking red wine all night in her flat. In the morning, she noticed an enormous <u>stain</u> on the carpet.
HF-LP	(a) Johnny enjoyed his first day at primary school. There was one particular <u>story</u> he liked about a tiger.
	(b) The child ran home with something he had taken from the garden. It was a small <u>stone</u> which had come from the gravel path.
HF-HP	(a) Johnny liked his father to read to him before bedtime. There was one particular <u>story</u> he liked about a tiger.
	(b) I could feel something in my shoe which dug into my heel. It was a small <u>stone</u> which had come from the gravel path.

Note: Target words are underlined. LF = low frequency; HF = high frequency; LP = low predictability; HP = high predictability.

Plausibility ratings were obtained from typically-developed adult native speakers of English ($N = 10$) and of German ($N = 10$). Participants were asked to rate the plausibility of target words on a scale from 1 (highly implausible) to 7 (highly plausible), with 4 indicating neutrality (neither plausible nor implausible). We compared lower- and higher-predictability conditions across the two material sets. The mean plausibility rating of LP targets was 4.73 ($SD = 1.91$) for our materials and 3.63 ($SD = 2.23$) for Dambacher et al.'s materials. A Mann-Whitney U test revealed that the materials in the current study were rated as significantly more plausible than those employed by Dambacher et al. [$U = 188458.5$, $z = -10.064$, $p < .001$]. For HP targets, the mean plausibility rating was 6.30 ($SD = 0.97$) for our materials and 5.39 ($SD = 1.68$) for Dambacher et al.'s materials. As before, the analysis demonstrated that our HP materials were rated as significantly more plausible than those of Dambacher et al. [$U = 36889.0$, $z = -7.337$, $p < .001$].

For the EEG experiment, two sets of materials were created. Only one version of each passage (LP or HP) was included in each set. Each set had a similar profile in terms of word length, frequency, and predictability characteristics of target words.² With a total of 248 target words, each of the two sets of materials comprised equal numbers of LF-LP, LF-HP, HF-LP, and HF-HP items ($N = 62$).

Apparatus

Participants were tested individually in an electrically shielded booth with low level ambient light. Experimental Run Time System (ERTS) software was used to control stimulus presentation (Dutta, 1995). Visual stimuli were presented on a 21" 1100 MB Samsung SyncMaster screen with a resolution of 1280 × 960 pixels and a refresh rate of 60 Hz. Participants were seated at a viewing distance of approximately 65 cm from the monitor, maintained throughout the experiment by means of a chin rest. Stimuli were presented centrally in Helvetica font (14-point for the first sentence; 16-point for the word-by-word presentation of the second sentence) in white letters on a black background.

Procedure

Participants were first informed about the nature of electrophysiological recording and were given specific task instructions. They were told that they would be reading several two-sentence passages of text, that each passage was like a very short story, and that they should read normally for comprehension. They were

instructed that the first sentence would be presented in full and that the second would be presented word-by-word. They were asked to maintain fixation at the centre of the screen during the word-by-word presentations. Following electrode application, participants sat facing the computer screen with their heads stabilised via a chin rest.

Each trial began with a central red fixation cross presented for 500 ms, signalling the start of a new passage, replaced by a white fixation cross for another 500 ms. The first, context sentence was then presented on the screen for a minimum duration of 1500 ms. When participants had finished reading the sentence, they pressed the spacebar on the computer keyboard which initiated the word-by-word presentation of the second, target sentence. Once the spacebar was pressed, a blank screen was presented for 500 ms, followed by a central white fixation cross for 500 ms. Then each word of the second sentence was displayed centrally for 267 ms, with a 33 ms blank interval between successive word presentations. After the final word of the second sentence, the screen remained blank for 1000 ms before the next trial began.

Participants were first presented with a practice block of two trials to become accustomed to the procedure and presentation of materials. All 248 experimental trials were then presented in a different random order for each participant, divided into 8 blocks of 31 trials with self-paced rest periods. The entire experiment including participant preparation lasted approximately 1.5 h.

Electrophysiological measures

EEG activity was recorded continuously from 72 Ag-AgCl electrodes using a BIOSEMI Active-Two amplifier system. Midline electrodes were Fpz, AFz, Fz, FCz, Cz, CPz, Pz, POz, Oz, and Iz; left hemisphere electrodes were IO1, Fp1, AF3, AF7, F1, F3, F5, F7, F9, FC1, FC3, FC5, FT7, C1, C3, C5, M1, T7, CP1, CP3, CP5, TP7, P1, P3, P5, P7, PO3, PO7, O1, two nonstandard positions PO9' located at 33% and O9' located at 66% of the M1-Iz distance, and the right hemisphere electrodes were the homologous sites. The sampling rate for the EEG and electrooculogram (EOG) recordings was 256 Hz. Off-line, all EEG channels were recalculated to an average reference and high-pass filtered (0.1 Hz, 12 dB/oct). Next, as described in more detail by Dudschig, Mackenzie, Strozyk, Kaup, and Leuthold (2016), a procedure similar to that of Nolan, Whelan, and Reilly (2010) was used to remove (ocular) artifacts and to correct EEG data. More specifically, a predefined z-score threshold of ± 3 was used to identify outliers relating to channels, epochs, independent components, and single-channels in single-epochs.

After removing epochs containing extreme values in single electrodes (e.g. amplifier blockings, values $> \pm 1000 \mu\text{V}$ in any electrode) and trials containing values exceeding $\pm 75 \mu\text{V}$ in multiple adjacent electrodes unrelated to eye movements, z-scored variance measures were calculated for all electrodes. Noisy EEG electrodes (z-score $> \pm 3$) were removed if their activity was uncorrelated to EOG activity and this "cleaned" EEG data set was subjected to a spatial independent components analysis (ICA) based on the infomax algorithm (Bell & Sejnowski, 1995). ICA components representing ocular activity (blinks and horizontal eye movements) were automatically identified using z-scored measures of the absolute correlation between the ICA component and the recorded hEOG and vEOG activity, respectively, and confirmed by visual inspection. Then, previously removed noisy channels were interpolated in the ICA-cleaned EEG data set using the average EEG activity of adjacent uncontaminated channels within a specified distance (4 cm, ~ 3 –4 neighbours per electrode). This ensured a full electrode array for each participant. Following artifact rejection and correction, there remained on average 58.8 trials (out of 62) per condition (range: 49–61; median = 60).

Data analysis

All EEG/ERP analysis was performed using available MATLAB toolboxes (EEGLAB: Delorme & Makeig, 2004; FieldTrip: Oostenveld, Fries, Maris, & Schoffelen, 2011) and custom MATLAB scripts. The analysis epoch started 1000 ms prior to the onset of the critical word and lasted until 1500 ms after it, resulting in a total epoch duration of 2500 ms. Off-line, all EEG channels were recalculated to an average mastoid reference. For artifact-free trials, the signal at each electrode site was averaged separately for each experimental condition, time-locked to the onset of the critical word, low-pass filtered (30 Hz, 36 dB/oct), and aligned to a 200 ms baseline prior to the onset of the critical word. In line with the analysis procedures of ERPs established in previous research (e.g. Dambacher et al., 2009; Hauk & Pulvermüller, 2004; Scott et al., 2009), mean amplitudes of specific ERP deflections were measured for the following time intervals: 50–80 ms, 80–120 ms (P1); 160–200 ms (N1), 200–300 ms (N2), and from 350 to 550 ms (N400). In addition, aligned to a 200 ms baseline interval preceding the pre-critical word (–500 to –300 ms), we determined mean ERP amplitude during the 200 ms interval immediately before the onset of the critical word (–200–0 ms).

Similar to Dambacher et al. (2009), we conducted an omnibus analysis of variance (ANOVA) with predefined regions of interest (ROIs), with electrodes pooled in nine

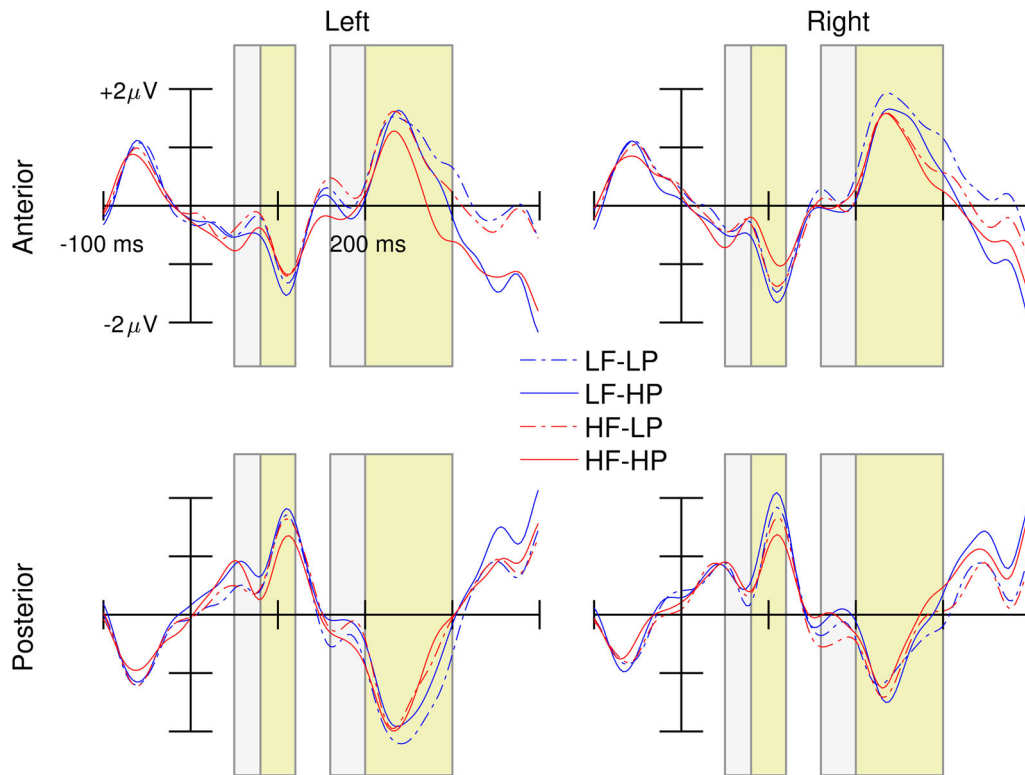


Figure 2. Average ERP waveforms for experimental conditions in left and right anterior and posterior ROIs. Time windows include 50–80 ms, 80–120 ms (P1), 160–200 ms (N1), and 200–300 ms (N2).

for anterior to posterior ROIs showed no reliable context effects [all $F_s < 2.30$, $p_s > .14$].

50–80 ms

The ANOVA revealed a reliable Predictability \times AntPost interaction [$F(2,58) = 5.82$, $p = .019$, $\eta_p^2 = 0.17$], and a Predictability \times Hemisphere \times AntPost interaction [$F(4,116) = 2.62$, $p = .039$, $\eta_p^2 = 0.08$]. Further testing showed reliable Predictability effects (LP vs. HP) for the left-anterior ROI (-0.27 vs. $-0.54 \mu V$) [$F(1,29) = 8.08$, $p = .008$], the mid-anterior ROI (-0.38 vs. $-0.71 \mu V$) [$F(1,29) = 7.51$, $p = .01$], and for the left-posterior ROI (0.43 vs. $0.71 \mu V$) [$F(1,29) = 4.85$, $p = .036$]. No other effects of experimental factors with topographic variables were significant [all $F_s < 2.08$, $p_s > .13$].

P1 (80–120 ms)

The ANOVA showed a Frequency \times AntPost interaction [$F(2,58) = 7.73$, $p = .006$, $\eta_p^2 = 0.21$], due to a more negative-going mean P1 amplitude for LF than HF words over anterior ROIs (-1.14 vs. $-0.90 \mu V$) [$F(1,29) = 7.04$, $p = .013$] and a more positive-going mean P1 amplitude over posterior ROIs (0.78 vs. $0.52 \mu V$) [$F(1,29) = 8.33$, $p = .007$], but no difference over central ROIs (-0.58 vs. $-0.53 \mu V$) [$F(1,29) = 2.45$, $p = .13$].

Crucially, the Frequency \times Predictability \times AntPost interaction was significant [$F(2,58) = 8.93$, $p = .003$, $\eta_p^2 = 0.24$]. The Frequency effect (LF vs. HF) in mean P1 amplitude was present only for HP items over anterior ROIs (-1.28 vs. $-0.83 \mu V$) [$F(1,29) = 10.71$, $p = .003$], central ROIs (-0.63 vs. $-0.47 \mu V$) [$F(1,29) = 7.43$, $p = .011$], and posterior ROIs (0.90 vs. $0.45 \mu V$) [$F(1,29) = 14.13$, $p < .001$], but was absent for the LP items [all $F_s < 1$]. Further tests indicated no reliable Predictability effect in P1 amplitude either for HF words [all $F_s < 1.55$, $p_s > .21$] or for LF words [all $F_s < 3.73$, $p_s > .06$].

N1 (160–200 ms)

During this time interval, the ANOVA showed a trend for the Predictability \times Hemisphere \times AntPost interaction [$F(4,116) = 2.34$, $p = .059$, $\eta_p^2 = 0.07$], due to a more negative-going amplitude for LP than HP items over the midline-posterior ROI (0.38 vs. $0.79 \mu V$) [$F(1,29) = 7.46$, $p = .011$] and over the right-posterior ROI (-0.35 vs. $-0.05 \mu V$) [$F(1,29) = 5.23$, $p = .03$].

There was a significant Frequency \times Predictability \times Hemisphere interaction [$F(2,58) = 5.36$, $p = .007$, $\eta_p^2 = 0.16$]. A Frequency effect (LF vs. HF) in mean N1 amplitude was present for HP items over left-hemispheric ROIs (-0.06 vs. $-0.22 \mu V$) [$F(1,29) = 4.29$, $p = .047$] and

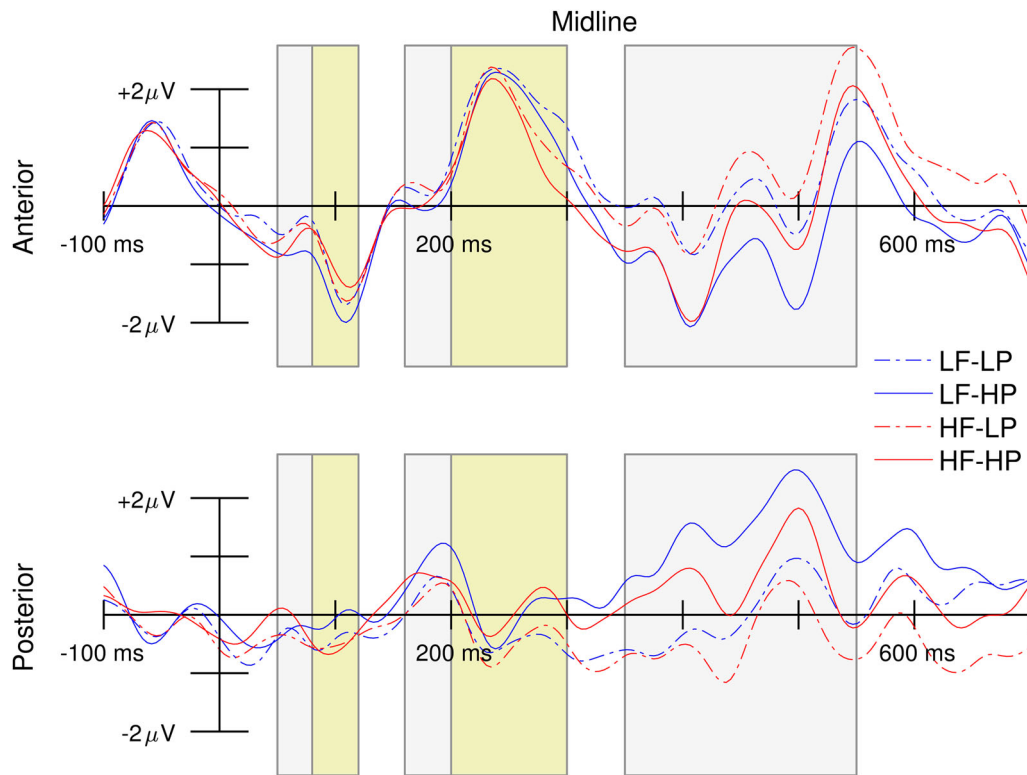


Figure 3. Average ERP waveforms for experimental conditions in midline anterior and posterior ROIs. The time window of interest is 350–550 ms (N400), although the earlier windows used in Figure 2 are also highlighted.

for LP items over left-hemispheric ROIs (-0.12 vs. $0.12 \mu\text{V}$) [$F(1,29) = 5.70$, $p = .024$], as well as over right-hemispheric ROIs (0.01 vs. -0.21) [$F(1,29) = 4.40$, $p = .045$]. There were no reliable Frequency effects over the other ROIs [all $F_s < 1.73$, $p_s > .19$]. Regarding the Predictability effect, further tests indicated that for HF words there was a more positive-going N1 amplitude for LP than HP items over left-hemispheric ROIs (0.12 vs. $-0.22 \mu\text{V}$) [$F(1,29) = 11.21$, $p = .002$], but a more negative-going amplitude over right-hemispheric ROIs (-0.21 vs. $0.05 \mu\text{V}$) [$F(1,29) = 6.53$, $p = .016$]. There was no reliable Predictability effect for HF words over

midline ROIs [$F(1,29) = 1.79$, $p = 0.19$], nor for LF words over any ROI [all $F_s < 1$].

N2 (200–300 ms)

The ANOVA showed a trend for the Frequency \times Hemisphere interaction [$F(2,58) = 3.28$, $p = .055$, $\eta_p^2 = 0.10$], indicating an enhanced positivity for LF than HF words over midline ROIs (1.02 vs. $0.82 \mu\text{V}$) [$F(1,29) = 15.84$, $p < .001$], but not over left- and right-hemispheric ROIs [all $F_s(1,29) < 1.68$, $p_s > .20$]. The Frequency \times AntPost interaction was significant [$F(2,58) = 6.58$, $p = .009$, $\eta_p^2 = 0.18$] due to an enhanced positivity for LF than HF

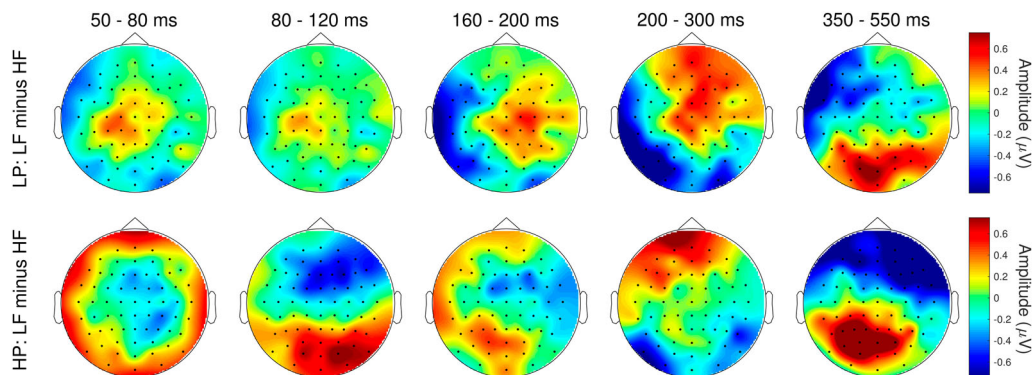


Figure 4. Scalp topographies of the mean Frequency difference (LP: LF minus HF; HP: LF minus HF) in ERP amplitudes across all time windows.

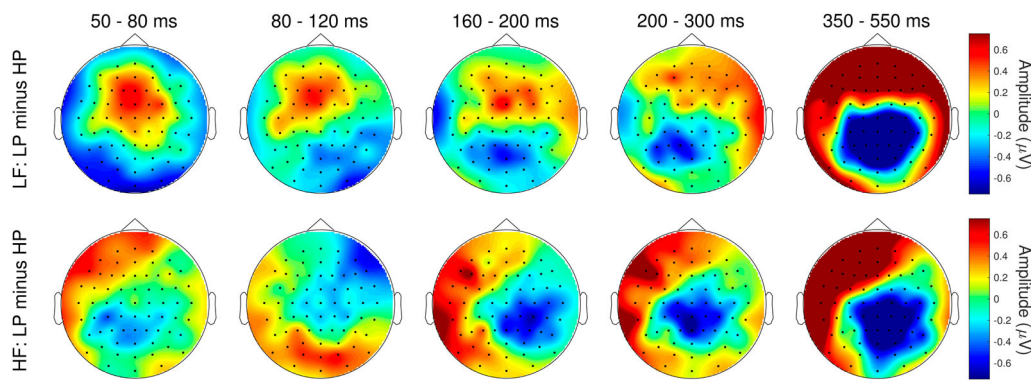


Figure 5. Scalp topographies of the mean Predictability difference (LF: LP minus HP; HF: LP minus HP) in ERP amplitudes across all time windows.

words over anterior ROIs (1.32 vs. 0.99 μV) [$F(1,29) = 11.32, p = .002$]. The Predictability \times AntPost interaction [$F(2,58) = 6.09, p = .016, \eta_p^2 = 0.17$] and the Predictability \times Hemisphere \times AntPost interaction were significant [$F(4,116) = 3.14, p = .017, \eta_p^2 = 0.10$]. The latter interaction indicated a more positive-going amplitude for LP than HP items over the left-anterior ROI (1.01 vs. 0.62 μV) [$F(1,29) = 7.39, p = .011$], but a more negative-going amplitude over the midline-posterior ROI (−0.37 vs. 0.12 μV) [$F(1,29) = 18.91, p < .001$].

Finally, there was a significant Frequency \times Predictability \times Hemisphere interaction [$F(2,58) = 4.47, p = .016, \eta_p^2 = 0.13$]. Further tests indicated that a Frequency effect (LF vs. HF) was present for LP items over left-hemispheric ROIs (−0.29 vs. −0.10 μV) [$F(1,29) = 6.66, p = .015$], over midline ROIs (0.98 vs. 0.70 μV) [$F(1,29) = 9.46, p = .005$], and over right hemispheric ROIs (0.35 vs. 0.17 μV) [$F(1,29) = 4.27, p = .048$]. There were no reliable Frequency effects for HP items [all $F_s < 3.37, p_s > .07$]. A Predictability effect for HF words was indicated by a more positive-going ERP amplitude for LP than HP items over left-hemispheric ROIs (−0.10 vs. −0.35 μV) [$F(1,29) = 9.68, p = .004$] and a more negative-going one over midline ROIs (0.70 vs. 0.93 μV) [$F(1,29) = 6.44, p = .017$], but not over right-hemispheric ROIs (0.17 vs. 0.26 μV) [$F(1,29) = 1.41, p = .25$]. There were no reliable Predictability effects for LF words [all $F_s < 2.23, p_s > .14$].

N400 (350–550 ms)

In the N400 time interval, the ANOVA revealed a significant Frequency \times AntPost interaction [$F(2,58) = 23.11, p < .001, \eta_p^2 = 0.44$], reflecting an enhanced negativity for LF than HF words over anterior ROIs (−0.68 vs. −0.19 μV) [$F(1,29) = 17.86, p < .001$] but an enhanced positivity over posterior ROIs (0.76 vs. 0.20 μV) [$F(1,29) = 32.97, p < .001$]. There were also significant interactions of Predictability \times Hemisphere [$F(2,58) = 20.91, p < .001, \eta_p^2 = 0.42$], Predictability \times AntPost [$F(2,58) = 57.51,$

$p < .001, \eta_p^2 = 0.66$], and Predictability \times Hemisphere \times AntPost [$F(4,116) = 16.37, p < .001, \eta_p^2 = 0.36$]. The latter interaction indicated over anterior ROIs a more positive-going ERP amplitude for LP than HP items that was most pronounced over the left hemisphere (−0.01 vs. −1.10 μV) [$F(1,29) = 49.40, p < .001$]. By contrast, over central and posterior ROIs, there was an enhanced negativity for LP than HP items and this Predictability effect was strongest over midline-central ROIs (−0.27 vs. 0.34 μV) [$F(1,29) = 34.70, p < .001$], and midline-posterior ROIs (−0.21 vs. 1.06 μV) [$F(1,29) = 87.60, p < .001$].

The Frequency \times Predictability \times Hemisphere interaction was significant [$F(2,58) = 4.52, p = .022, \eta_p^2 = 0.13$]. Further tests indicated that a Frequency effect (LF vs. HF) was present for HP items over right hemispheric ROIs (−0.31 vs. −0.05 μV) [$F(1,29) = 11.14, p = .002$]. There were no reliable Frequency effects over the other ROIs and also not for LP items [all $F_s < 3.32, p_s > .078$]. Regarding the Predictability effect, for HF words there was a more positive-going ERP amplitude for LP than HP items over left-hemispheric ROIs (0.40 vs. −0.15 μV) [$F(1,29) = 27.30, p < .001$] but a more negative-going one over midline ROIs (−0.06 vs. 0.20 μV) [$F(1,29) = 7.05, p = .013$] and over right-hemispheric ROIs (−0.33 vs. −0.05 μV) [$F(1,29) = 10.30, p = .003$]. There were reliable Predictability effects for LF words over left-hemispheric ROIs (0.28 vs. 0.03 μV) [$F(1,29) = 5.66, p = .024$] and over midline ROIs (−0.05 vs. 0.30 μV) [$F(1,29) = 14.99, p < .001$], but not over right-hemispheric ROIs (−0.25 vs. −0.31 μV) [$F < 1$].

Finally, there was a trend for the Frequency \times Predictability \times AntPost interaction [$F(2,58) = 3.22, p = .072, \eta_p^2 = 0.10$]. For HP items, there was an enhanced negativity for LF than HF words over anterior ROIs (−1.21 vs. −0.56 μV) [$F(1,29) = 19.04, p < .001$] and an enhanced positivity over posterior ROIs (1.20 vs. 0.49 μV) [$F(1,29) = 39.41, p < .001$]. For LP items, there was an enhanced positivity for LF than HF words over posterior ROIs only (0.33 vs.

$-0.08 \mu\text{V}$ [$F(1,29) = 10.52$, $p = .003$]. There were reliable Predictability effects for both LF and HF words over anterior to posterior ROIs [all $F_s(1,29) > 7.50$, $p_s \leq .01$].

Summary of findings

Frequency

Significant word frequency effects were demonstrated in the P1, N2, and N400 components. In the P1, LF words, in comparison to HF words, demonstrated enhanced negativity in anterior ROIs and enhanced positivity in posterior ROIs. In the N2, LF words showed enhanced positivity over HF words in midline and anterior ROIs. In the N400, LF words demonstrated enhanced negativity in anterior ROIs and enhanced positivity in posterior ROIs.

Predictability

Significant contextual predictability effects occurred in all time windows except the P1. In the 50–80 ms window, LP words were more positive-going than HP words over left- and midline-anterior ROIs, but more negative-going than HP words over the left-posterior ROI. In the N1, LP words were more negative-going than HP words over midline- and right-posterior ROIs. In the N2, LP words were more positive-going than HP words in the left-anterior ROI, but were more negative-going in the midline-posterior ROI. In the N400, there were widespread predictability effects, with the most pronounced differences emerging in the left-anterior ROI, with LP words more positive-going than HP words, and in midline-central and midline-posterior ROIs, with LP words more negative-going.

Frequency \times Predictability

Frequency and predictability interacted in all but the earliest time window. In the P1, frequency effects emerged selectively for HP words in anterior and central ROIs (enhanced negativity to LF words) as well as posterior ROIs (enhanced positivity to LF words). There were no frequency effects for LP words, nor predictability effects for LF or HF words. In the N1, frequency effects occurred in LP and HP words. For LP words, LF words were more negative-going over left-hemispheric ROIs and more positive-going over right-hemispheric ROIs. For HP words, in contrast, LF words were more positive-going over left-hemispheric ROIs. N1 predictability effects only occurred in HF words, with greater positivity to LP than HP words over left-hemispheric ROIs and greater negativity to LP words over right-hemispheric ROIs. In the N2, frequency effects emerged selectively for LP words in left-hemispheric ROIs (greater negativity to LF than HF words) and in midline and right-

hemispheric ROIs (greater positivity to LF than HF words). N2 predictability effects emerged selectively to HF words in left-hemispheric ROIs (more positive-going in LP than HP words) and in midline ROIs (more negative-going in LP than HP words). In the N400, in HP words, frequency effects occurred in right-hemispheric and anterior ROIs (greater negativity to LF than HF words) and in posterior ROIs (greater positivity to LF than HF words). LP words only demonstrated a frequency effect in posterior ROIs (greater positivity to LF words). N400 predictability effects were widespread in both LF and HF words, demonstrating greater positivity to LP than HP words in left-hemispheric ROIs and greater negativity to LP than HP words in midline ROIs and right-hemispheric ROIs (HF only). Predictability effects were also reliable for both LF and HF words across anterior, central, and posterior ROIs, with greater positivity to LP than HP words over anterior ROIs, but greater negativity to LP than HP words over central and posterior ROIs.

Discussion

Our study examined the relative timing of word frequency and contextual predictability effects in reading using electrophysiological recordings. Our approach is unique in that we implemented several methodological procedures to assure a systematic investigation of such effects. In general, past ERP studies that have examined word frequency and predictability effects have typically used more than one of the following: relatively short contexts; biasing contexts containing semantic primes that are proximal to the target (e.g. Kretzschmar et al., 2015; Penolazzi et al., 2007); potentially anomalous contexts to represent “low predictability” conditions (e.g. Dambacher et al., 2012); sentence-final targets; and/or a slow presentation rate. Such procedures may not promote discourse processing associated with normal reading. Our methodology addressed these concerns in several ways. We employed a 2 (Frequency: LF, HF) \times 2 (Predictability: LP, HP) design using a high number of items per condition ($N = 62$). Contexts were relatively long and independent from the target sentence. As neutral and biasing contexts were tailor-made for each target word, LP targets were not anomalous. Target words were not sentence-final, but were embedded within the discourse. Target position was not confounded with its predictability. Finally, a rapid, reading-like rate of 300 ms/word was used for target sentence presentation.

ERPs to target words across conditions were analysed across five time windows: 50–80 ms, 80–120 ms (P1), 160–200 ms (N1), 200–300 ms (N2), and 350–550 ms (N400). Analyses of the data suggest a complex pattern

of results. We will first summarise our results in each successive time window, relating our findings to those of previous studies.

The earliest time window of 50–80 ms revealed a predictability effect with a more positive-going waveform to LP than HP words over left- and midline-anterior ROIs, and a more negative-going waveform to LP than HP words over the left-posterior ROI, which cannot be attributed to ERP differences arising in the baseline interval preceding the critical word. Dambacher et al. (2009) is the only other study to our knowledge that has also reported such an early (50–90 ms) predictability effect. However, the topography and direction of their effect differed, with the LP condition more negative than HP in right- and midline-anterior sites, but more positive than HP in left-posterior sites. It is interesting to note that in an auditory sentence processing study, van Berkum, Brown, Zwitserlood, Kooijman, and Hagoort (2005) report a larger anterior as well as right-posterior positivity to a prediction inconsistent than prediction consistent word between 50 and 250 ms. We are unable to offer an explanation for these mixed patterns of early ERP predictability effects aside from procedural differences.

In the P1 (80–120 ms), word frequency effects emerged, showing greater positivity in posterior sites, with reversed polarity in anterior sites. This is in accord with Hauk et al. (2006) who found a larger P1 to LF than HF words over left posterior sites. This frequency effect was modulated by predictability, with follow-up tests showing that the effect was limited to HP words. LF-HP words showed greater positivity than HF-HP words over posterior ROIs, but greater negativity in central and anterior ROIs. This is a novel finding demonstrating early interactive processing. In a similar time window (110–130 ms), Penolazzi et al. (2007) had found that both frequency and predictability interacted separately with word length, however, these factors did not interact with each other. It is possible that their study had insufficient power given their lower number of participants in comparison to that of the current study (17 vs. 30, respectively).

In the N1 (160–200 ms), the frequency-predictability interaction again emerged, but with a different pattern of effects. There were frequency effects for both HP and LP conditions. Over left-hemispheric ROIs, ERPs to LF words were more positive than HF words in the HP condition, but were more negative than HF words in the LP condition. In the LP condition, LF words were additionally more positive than HF words in right-hemispheric ROIs. Sereno et al. (2003) also found a frequency-predictability interaction in a similar time window (132–192 ms). They found that although ERPs were more negative to LF

than HF words in both LP and HP conditions, the effect was of smaller magnitude in the HP condition. Dambacher et al. (2012), using a 280 ms SOA, also showed a frequency-predictability interaction in an early N1 (135–155 ms) and a significant effect of frequency in a later N1 (190–260 ms). In the early N1, frequency effects for HP words were more negative over posterior channels and more positive over anterior ones.

In the N2 (200–300 ms), there was a frequency effect over both midline and anterior ROIs with a greater positivity to LF than HF words. As in the N1, the frequency effect was modulated by predictability in interaction with hemisphere. However, only LP conditions demonstrated frequency effects in terms of a more negative-going ERP to LF than HF words over left-hemispheric ROIs, but a more positive-going ERP over midline- and right-hemispheric ROIs.

In the N400 (350–550 ms), there was a frequency effect that emerged as a larger negative-going amplitude to LF than HF words over anterior ROIs and this frequency effect was reversed in polarity over posterior sites, possibly indicating a P300-like effect (cf. Polich & Donchin, 1988). Crucially, this effect was again modulated by predictability, as it was present over anterior ROIs only for the HP but not the LP condition, whereas both HP and LP conditions showed a reliable frequency effect over posterior ROIs. In addition, there was a frequency effect over left-hemispheric ROIs limited to the HP condition. Previous studies have demonstrated in the N400 that frequency effects are present only for LP but not for HP conditions (Dambacher et al., 2006; van Petten & Kutas, 1990), that frequency and predictability produce additive effects (Penolazzi et al., 2007), or that only predictability but not frequency effects occur (Dambacher et al., 2012; Kretzschmar et al., 2015).

Finally, there was also a predictability effect that was present over posterior-midline ROIs in the three successive time windows from 160 to 550 ms. This effect was characterised by a greater negative-going ERP for LP than HP conditions. Prior studies have also demonstrated significant predictability effects, with greater negativity for LP than HP conditions, with a similar topography: Dambacher et al. (2012), using a 280 ms SOA, found a predictability effect from 300 to 500 ms post-stimulus; Kretzschmar et al. (2015) showed sustained effects of predictability in successive 50 ms time windows, from 150 to 650 ms after fixation onset.

Taken together, our findings are generally supportive of other studies that have examined word frequency and contextual predictability effects in reading. Undoubtedly, there are certain discrepancies in terms of the topography and/or the polarity of effects. It is possible that the procedure we implemented, in particular, the rapid

presentation rate, could have given rise to some of the differential findings. For example, if contextual benefits, broadly construed, were already expressed on the pre-target word, then the later-onset components associated with the pre-target word would be occurring during the same time interval as the earlier components associated with the target word. If this were the case, it would suggest an even earlier temporal onset of predictive processing that potentially may be additionally sensitive to the frequency of the upcoming word. Dimigen et al. (2011), however, observed that component overlap can be an issue, but only if the amount of overlap systematically differs between conditions. Target sentences in our LP and HP conditions were identical in terms of their content and word presentation rate.

Perhaps the most compelling aspect of our findings is the confirmation of an early time course of effects in an experimental situation akin to reading within a natural discourse context. The onsets of the main effects of word frequency and contextual predictability replicated those reported, respectively, by Hauk et al. (2006) in the P1 and by Dambacher et al. (2009) in an interval preceding the P1. We additionally demonstrated an interaction of frequency and predictability with a P1 onset that was sustained through the N400. These electrophysiological results can be used to constrain computational models of written language processing (for a review, see Barber & Kutas, 2007). There is growing evidence for rapid neural sensitivity and response to expectation in visual perception, in particular, via an early top-down influence from orbitofrontal cortex (Hofmann et al., 2014; Trapp & Bar, 2015). Our early time frame and topography of effects suggest that context can influence activation of low-level visual features, a component central to models of language processing (Granger & Holcomb, 2009; Hofmann & Jacobs, 2014; Price & Devlin, 2011). Our findings of an early and robust frequency-predictability interaction provide strong evidence for interactive processing during lexical access in reading.

Notes

1. Our mean HP Cloze values were .61, which is in line with HP conditions of eye movement reading studies that have examined frequency-predictability effects (see, e.g., Sereno et al., 2018). Moreover, the 95% confidence intervals (CIs) of our four conditions demonstrated clear boundaries between LP and HP classifications across levels of Frequency: LF-LP CI [0.00, 0.07]; LF-HP CI [0.50, 0.72]; HF-LP CI [0.01, 0.11]; and HF-HP CI [0.50, 0.74]. Nevertheless, the mean Cloze values of HP conditions were higher in both the Dambacher et al. (2012) and Kretzschmar et al. (2015) studies (0.835 and

0.785, respectively). However, certain methodological issues should be considered. In Dambacher et al.'s Cloze procedure, participants were allowed three guesses of the upcoming target word (typically, only one is given), functionally increasing the likelihood that the target would be guessed. In Kretzschmar et al., context was not provided in a separate sentence, but was limited to the first few words of the sentence. As such, although the mean Cloze values for HP conditions were relatively high, there was an increased risk of context operating via intralexical priming and not higher-level discourse processes (e.g., in Table 1, the words *Sunday*, *nun*, and *pray* closely precede the target *church*).

2. We performed analyses on the two alternative sets of materials in order to confirm whether there were any differences in terms of target length, frequency, and predictability, as well as number of words preceding the target, and number of characters in the first context sentence. No evidence of any differences between the two sets of materials was found on any of these dimensions [all $F_s < 1$].

Disclosure statement

No potential conflict of interest was reported by the authors.

Funding

This work was supported by an Economic and Social Research Council [grant number RES-062-23-1900].

References

- Barber, H. A., & Kutas, M. (2007). Interplay between computational models and cognitive electrophysiology in visual word recognition. *Brain Research Reviews*, 53(1), 98–123.
- Bell, A. J., & Sejnowski, T. J. (1995). An information-maximization approach to blind separation and blind deconvolution. *Neural Computation*, 7(6), 1129–1159.
- Dambacher, M., Dimigen, O., Braun, M., Wille, K., Jacobs, A. M., & Kliegl, R. (2012). Stimulus onset asynchrony and the timeline of word recognition: Event-related potentials during sentence reading. *Neuropsychologia*, 50(8), 1852–1870.
- Dambacher, M., Kliegl, R., Hofmann, M., & Jacobs, A. M. (2006). Frequency and predictability effects on event-related potentials during reading. *Brain Research*, 1084(1), 89–103.
- Dambacher, M., Rolfs, M., Gollner, K., Kliegl, R., & Jacobs, A. M. (2009). Event-related potentials reveal rapid verification of predicted visual input. *PLoS One*, 4, e5047.
- Davies, M. (2004). BYU-BNC (based on the British National Corpus from Oxford University Press). Retrieved from <http://corpus.byu.edu/bnc/>
- DeLong, K. A., Urbach, T. P., & Kutas, M. (2005). Probabilistic word pre-activation during language comprehension inferred from electrical brain activity. *Nature Neuroscience*, 8(8), 1117–1121.
- Delorme, A., & Makeig, S. (2004). EEGLAB: An open source toolbox for analysis of single-trial EEG dynamics including

- independent component analysis. *Journal of Neuroscience Methods*, 134(1), 9–21.
- Dimigen, O., Sommer, W., Hohlfield, A., Jacobs, A. M., & Kliegl, R. (2011). Coregistration of eye movements and EEG in natural reading: Analyses and review. *Journal of Experimental Psychology: General*, 140(4), 552–572.
- Dudschig, C., Mackenzie, I. G., Strozyk, J., Kaup, B., & Leuthold, H. (2016). The sounds of sentences: Differentiating the influence of physical sound, sound imagery, and linguistically implied sounds on physical sound processing. *Cognitive, Affective, & Behavioral Neuroscience*, 16(5), 940–961.
- Dutta, A. (1995). Experimental run time system: Software for developing and running experiments on IBM-compatible PCs. *Behavior Research Methods, Instruments, and Computers*, 27(4), 516–519.
- Fodor, J. A. (1983). *The modularity of mind*. Cambridge, MA: MIT Press.
- Forster, K. I. (1979). Levels of processing and the structure of the language processor. In W. E. Cooper & E. Walker (Eds.), *Sentence processing: Psycholinguistic essays presented to Merrill Garrett* (pp. 27–85). Hillsdale, NJ: Erlbaum.
- Friedman, D., Simson, R., Ritter, W., & Rapin, I. (1975). The late positive component (P300) and information processing in sentences. *Electroencephalography and Clinical Neurophysiology*, 38(3), 255–262.
- Gold, B. T., Balota, D. A., Jones, S. J., Powell, D. K., Smith, C. D., & Andersen, A. H. (2006). Dissociation of automatic and strategic lexical-semantics: Functional magnetic resonance imaging evidence for differing roles of multiple frontotemporal regions. *Journal of Neuroscience*, 26(24), 6523–6532.
- Grainger, J., & Holcomb, P. J. (2009). Watching the word go by: On the time-course of component processes in visual word recognition. *Language and Linguistics Compass*, 3(1), 128–156.
- Hagoort, P. (2003). Interplay between syntax and semantics during sentence comprehension: ERP effects of combining syntactic and semantic violations. *Journal of Cognitive Neuroscience*, 15(6), 883–899.
- Hand, C. J., Miellet, S., O'Donnell, P. J., & Sereno, S. C. (2010). The frequency-predictability interaction in reading: It depends where you're coming from. *Journal of Experimental Psychology: Human Perception and Performance*, 36(5), 1294–1313.
- Hand, C. J., O'Donnell, P. J., & Sereno, S. C. (2012). Word-initial letters influence fixation durations during fluent reading. *Frontiers in Psychology: Language Sciences*, 3(85), 1–19. doi:10.3389/fpsyg.2012.00268
- Hauk, O., Davis, M. H., Ford, M., Pulvermüller, F., & Marslen-Wilson, W. D. (2006). The time course of visual word recognition as revealed by linear regression analysis of ERP data. *NeuroImage*, 30(4), 1383–1400.
- Hauk, O., & Pulvermüller, F. (2004). Effects of word length and frequency on the human event-related potential. *Clinical Neurophysiology*, 115(5), 1090–1103.
- Hillyard, S. A., & Münte, T. F. (1984). Selective attention to color and location: An analysis with event-related brain potentials. *Perception & Psychophysics*, 36(2), 185–198.
- Hofmann, M. J., Dambacher, M., Jacobs, A. M., Kliegl, R., Radach, R., Kuchinke, L., ... Hermann, M. J. (2014). Occipital and orbitofrontal hemodynamics during naturally paced reading: An fNIRS study. *NeuroImage*, 94, 193–202.
- Hofmann, M. J., & Jacobs, A. M. (2014). Interactive activation and competition models and semantic context: From behavioral to brain data. *Neuroscience and Biobehavioral Reviews*, 46(1), 85–104.
- Inhoff, A. W. (1984). Two stages of word processing during eye fixations in the reading of prose. *Journal of Verbal Learning and Verbal Behavior*, 23(5), 612–624.
- Kliegl, R., Grabner, E., Rolfs, M., & Engbert, R. (2004). Length, frequency, and predictability effects of words on eye movements in reading. *European Journal of Cognitive Psychology*, 16(1/2), 262–284.
- Kliegl, R., Nuthmann, A., & Engbert, R. (2006). Tracking the mind during reading: The influence of past, present, and future words on fixation durations. *Journal of Experimental Psychology: General*, 135(1), 12–35.
- Kretzschmar, R., Schlesewsky, M., & Staub, A. (2015). Dissociating word frequency and predictability effects in reading: Evidence from coregistration of eye movements and EEG. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 41(6), 1648–1662.
- Kutas, M., & Federmeier, K. D. (2011). Thirty years and counting: Finding meaning in the N400 component of the event-related brain potential (ERP). *Annual Review of Psychology*, 62, 621–647.
- Kutas, M., & Hillyard, S. A. (1984). Brain potentials during reading reflect word expectancy and semantic association. *Nature*, 307, 161–163.
- Leuthold, H., Kunkel, A., Mackenzie, I. G., & Filik, R. (2015). Online processing of moral transgressions: ERP evidence for spontaneous evaluation. *Social Cognitive and Affective Neuroscience*, 10(8), 1021–1029.
- McClelland, J. L. (1987). The case for interactionism in language processing. In M. Coltheart (Ed.), *Attention & performance XII: The psychology of reading* (pp. 1–36). London: Erlbaum.
- Miellet, S., Sparrow, L., & Sereno, S. C. (2007). Word frequency and predictability effects in reading French: An evaluation of the E-Z Reader model. *Psychonomic Bulletin & Review*, 14(4), 762–769.
- Murray, W. S., & Forster, K. I. (2004). Serial mechanisms in lexical access: The rank hypothesis. *Psychological Review*, 111(3), 721–756.
- Nolan, H., Whelan, R., & Reilly, R. B. (2010). FASTER: Fully automated statistical thresholding for EEG artifact rejection. *Journal of Neuroscience Methods*, 192(1), 152–162.
- Oostenveld, R., Fries, P., Maris, E., & Schoffelen, J. M. (2011). FieldTrip: Open source software for advanced analysis of MEG, EEG, and invasive electrophysiological data. *Computational Intelligence and Neuroscience*, 2011(156869). doi:10.1155/2011/156869
- Penolazzi, B., Hauk, O., & Pulvermüller, F. (2007). Early semantic context integration and lexical access as revealed by event-related brain potentials. *Biological Psychology*, 74(3), 374–388.
- Polich, J., & Donchin, E. (1988). P300 and the word frequency effect. *Electroencephalography and Clinical Neurophysiology*, 70(1), 33–45.
- Price, C. J., & Devlin, J. T. (2011). The interactive account of ventral occipitotemporal contributions to reading. *Trends in Cognitive Sciences*, 15(6), 246–253.
- Rayner, K., Ashby, J., Pollatsek, A., & Reichle, E. D. (2004). The effects of frequency and predictability on eye fixations in reading: Implications for the E-Z Reader model. *Journal of*

- Experimental Psychology: Human Perception and Performance*, 30(4), 720–732.
- Rayner, K., & Sereno, S. C. (1994). Eye movements in reading: Psycholinguistic studies. In M. A. Gernsbacher (Ed.), *Handbook of psycholinguistic research* (pp. 57–81). New York: Academic Press.
- Rugg, M. D. (1990). Event-related brain potentials dissociate repetition effects of high- and low-frequency words. *Memory & Cognition*, 18(4), 367–379.
- Scott, G. G., O'Donnell, P. J., Leuthold, H., & Sereno, S. C. (2009). Early emotion word processing: Evidence from event-related potentials. *Biological Psychology*, 80(1), 95–104.
- Sereno, S. C., Brewer, C. C., & O'Donnell, P. J. (2003). Context effects in word recognition: Evidence for early interactive processing. *Psychological Science*, 14(4), 328–333.
- Sereno, S. C., Hand, C. J., Shahid, A., Yao, B., & O'Donnell, P. J. (2018). Testing the limits of contextual constraint: Interactions with word frequency and parafoveal preview during fluent reading. *Quarterly Journal of Experimental Psychology*, 71(1), 302–313.
- Sereno, S. C., & Rayner, K. (2000). The when and where of reading in the brain. *Brain and Cognition*, 42(1), 78–81.
- Sereno, S. C., & Rayner, K. (2003). Measuring word recognition in reading: Eye movements and event-related potentials. *Trends in Cognitive Sciences*, 7(11), 489–493.
- Sereno, S. C., Rayner, K., & Posner, M. I. (1998). Establishing a time-line of word recognition: Evidence from eye movements and event-related potentials. *NeuroReport*, 9(10), 2195–2200.
- Stanovich, K. E., & West, R. F. (1983). On priming by a sentence context. *Journal of Experimental Psychology: General*, 112(1), 1–36.
- Sternberg, S. (1969). Memory-scanning: Mental processes revealed by reaction-time experiments. *American Scientist*, 57(4), 421–457.
- Trapp, S., & Bar, M. (2015). Prediction, context, and competition in visual recognition. *Annals of the New York Academy of Sciences*, 1339(1), 190–198.
- van Berkum, J. J. A., Brown, C. M., Zwitterlood, P., Kooijman, V., & Hagoort, P. (2005). Anticipating upcoming words in discourse: Evidence from ERPs and reading times. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 31(3), 443–467.
- van Berkum, J. J. A., Hagoort, P., & Brown, C. M. (1999). Semantic integration in sentences and discourse: Evidence from the N400. *Journal of Cognitive Neuroscience*, 11(6), 657–671.
- van Petten, C., & Kutas, M. (1990). Interactions between sentence context and word frequency in event-related brain potentials. *Memory & Cognition*, 18(4), 380–393.
- West, R. F., & Stanovich, K. E. (1982). Source of inhibition in experiments on the effect of sentence context on word recognition. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 8(5), 385–399.